

CLAIM SUMMARY DOCUMENT

Claims 1-18 (Canceled)

19. (New) A system for indexing textual content in any of a plurality of languages for searching purposes, comprising:

a tokenizer which separates a string of text into individual word tokens;

a stemmer which reduces the word tokens to grammatical stems by removing word endings which are associated with any one or more of the languages, without regard to whether the remaining stem is a recognized word in any combination of the plurality of languages; and

an index which stores the stems in an index.

20. (New) The system of claim 19 wherein the word endings which are removed are limited to only those endings which are associated with nouns.

21. (New) The system of claim 19 wherein a word ending is not removed if the resulting stem is less than a predetermined length.

22. (New) The system of claim 21 wherein said predetermined length is four characters.

Preliminary Amendment
Continuation of Application Serial No. 10,080,513
Attorney's Docket No. P2272USC2-942
Page 3

23. (New) The system of claim 19 wherein the stemmer reduces only once per word token.

24. (New) The system of claim 23 wherein said stemmer reduces by first examining each word token for the longest known endings, and examining the token for successively shorter endings until a known ending is identified in the word token and removed.

25. (New) The system of claim 19 wherein the stemmer and index disregard stopwords, wherein stopwords are words which occur with relatively high frequency in at least one of said languages and which are not also significant nouns in another one of said languages.

26. (New) A computer-readable medium containing a computer program for searching for documents that may contain text in any of a plurality of languages, wherein the computer program performs the steps of:

separating text in each document to be searched into individual word tokens;
reducing the word tokens to grammatical stems by removing word endings that are associated with any one or more of the languages, without regard to whether the remaining stem is a recognized word in any of the plurality of languages;

storing the stems in an index that identifies the documents in which words containing the stems appeared; parsing a query containing a string of text into individual word tokens; reducing the word tokens from the query to grammatical stems by removing word endings that are associated with any one or more of the languages, without regard to whether the remaining stem is a recognized word in any of the plurality of languages; searching the index for entries that match the stems obtained from the query; and displaying an identification of the documents that contained matching entries.

27. (New) The computer-readable medium of claim 26, wherein the computer program performs the step of:

displaying a matching entry along with the identification of the document in which it appears, wherein a stem is displayed together with an ending to present a full word to the user.

28. (New) The computer-readable medium of claim 26, wherein a stem is stored in the index together with the ending that was removed from a word token to form that stem, and an entry in the index that matches a stem from a query is displayed with the stored ending.

Preliminary Amendment
Continuation of Application Serial No. 10,080,513
Attorney's Docket No. P2272USC2-942
Page 5

29. (New) A system for searching for documents which may contain text in any of a plurality of languages, comprising:

a tokenizer which separates text in each document to be searched into individual word tokens;

a stemmer which reduces the word tokens to grammatical stems by removing word endings which are associated with any one or more of the languages, without regard to whether the remaining stem is a recognized word in any of the plurality of languages;

an indexer which stores the stems in an index and which identifies the documents in which words containing the stems appeared;

a search engine which parses a query containing a string of text into individual word tokens, reduces the word tokens from said query to grammatical stems by removing word endings which are associated with any one or more of the languages, without regard to whether the remaining stem is a recognized word in any of the plurality of languages, and searches the index for entries which match the stems obtained from said query; and

a display which displays an identification of the documents which contained matching entries.

30. (New) The system of claim 29 wherein the display displays a matching entry along with the identification of the document in which it appears, and wherein a stem is displayed together with an ending to present a full word to the user.

Preliminary Amendment
Continuation of Application Serial No. 10,080,513
Attorney's Docket No. P2272USC2-942
Page 6

31. (New) The system of claim 29 wherein a stem is stored in said index together with the ending that was removed from a word token to form that stem, and an entry in the index that matches a stem from a query is displayed with said stored ending.